

STATISTICAL MODEL ON STUDENT PERFORMANCE IN UTHM  
BY USING NON-PARAMETRIC, SEMI-PARAMETRIC AND  
PARAMETRIC SURVIVAL ANALYSIS

CHE AZMEEZA BINTI CHE HAMAT

A thesis submitted in  
fulfilment of the requirement for the award of the  
Degree of Master of Science

Faculty of Applied Sciences and Technology  
Universiti Tun Hussein Onn Malaysia

JUNE 2019

For my beloved parents Rahana and Che Hamat.



## ACKNOWLEDGEMENT

Thankful to God for the mercy, blessings and strength that were bestowed upon me in the quest to complete my thesis. This project bears the beautiful imprints of many peoples.

A token of gratitude to my beloved parents, Che Hamat Bin Che Harun and Rahana Binti Saamah for their love, support and inspiration during my study. Not to forget are my siblings, Azmiral, Azreen, Aznil and Ruslina; thank you very much for the support and motivation showered on me during this study.

My sincere gratitude and special thanks to my supervisor Professor Dr. Azme Bin Khamis for his insights, invaluable guidance, continuous assistance and encouragement throughout this research journey. I am especially indebted to him as a true mentor, the ultimate supervisor and motivator. I may not be able to make it this far without his perseverance, faith and patience. I am also grateful to my co-supervisor, Dr. Mohd Asrul Affendi Bin Abdullah for his guidance, opinion and motivation as well. I really appreciate the guidance from the initial to the final levels that enables me to develop thorough understanding of this research.

I am also sincerely thankful to the Research, Innovation, Commercialization & Consultancy Office (ORICC) Universiti Tun Hussein Onn Malaysia for providing the support and sponsors to do this project.

I am also indebted to many other people who have helped me directly or indirectly in this study. This thesis may not be successfully completed without the memorable involvement of my many lecturers, friends and others. I sincerely thank you.

## ABSTRACT

Student performance defined as students who are capable to success during their studies. This study explored the use of survival analysis to investigate the performance of Bachelor's degree students in Universiti Tun Hussein Onn Malaysia (UTHM). The data was collected from the Academic Management Office (AMO), UTHM. The main objective of this study is to estimate the survival rates of students with different entrance qualifications. The study also aim to identify the covariates that dominate the student performance, investigate the performance of Cox model based on the violation of the Proportional Hazard (PH) assumption, compare the model performance by using the survival and Accelerated Failure Time (AFT) models and estimate the time ratio (TR) of student performance in accordance to the selected best model. The survival analysis considered the survival approach such as the Kaplan-Meier (KM) method in the non-parametric method, Cox model in semi-parametric model and survival and AFT models in parametric model. The results revealed that students with STPM-entrance qualification had the highest survival rate compared to Diploma and Matriculation holders. The Cox model in the semi-parametric model identified the GPA, entrance qualification and course as the significant covariates to be included in the study. Faculty covariate was excluded since the  $p$ -value insignificant at 90% significance level. The result provided by the Cox model violated the PH assumptions. Then, the performance of the Cox model is less accurate. The invalidation performance of Cox model prompted the need to conduct other parametric survival and AFT models to produce more precise results. As a conclusion, the Log-normal AFT model is the best alternative model to estimate student performance in UTHM and other similar higher educational institution.

## ABSTRAK

Prestasi pelajar ditakrifkan sebagai kemampuan pelajar untuk cemerlang semasa pengajian. Kajian ini meneroka penggunaan analisis kemandirian untuk menyiasat keupayaan pelajar Ijazah Sarjana Muda di Universiti Tun Hussein Onn Malaysia (UTHM). Data dikumpulkan dari Pejabat Pengurusan Akademik (PPA), UTHM. Objektif utama kajian ini adalah untuk menganggarkan kadar kemandirian pelajar di universiti dengan kelayakan kemasukan yang berbeza. Kajian ini juga untuk mengenal pasti kovariat yang menguasai prestasi pelajar, menyiasat prestasi model Cox berdasarkan pelanggaran anggapan bahaya berkadaran (PH), membandingkan prestasi model dengan menggunakan model kemandirian parametrik dan masa kegagalan dipercepat (AFT). Akhir sekali menganggarkan nisbah masa (TR) prestasi pelajar mengikut model terbaik yang dipilih. Kajian ini menggunakan tiga kaedah iaitu kaedah Kaplan Meier (KM) dalam kaedah bukan parametrik, model Cox dalam separa parametrik dan model kemandirian parametrik dan model parametrik AFT dalam model parametrik untuk analisis kemandirian. Hasil kajian menunjukkan bahawa pelajar berkecualan STPM mempunyai kadar kelangsungan meneruskan pengajian yang paling tinggi berbanding pemegang Diploma dan Matrikulasi. Model Cox dalam model separa parametrik mengenal pasti GPA, kelayakan masuk dan kursus sebagai kovariat yang signifikan untuk dimasukkan ke dalam kajian ini manakala kovariat fakulti tidak dimasukkan kerana tidak memberikan nilai signifikan pada tahap 90%. Keputusan analisis model Cox tidak memenuhi anggapan PH. Hal ini menyebabkan prestasi yang dihasilkan model Cox kurang tepat. Ketidakstabilan prestasi model Cox mendorong kajian dijalankan menggunakan model kemandirian parametrik and model AFT parametrik untuk menghasilkan keputusan kajian yang lebih tepat. Kajian menyimpulkan bahawa model AFT Log-normal adalah model alternatif terbaik untuk menganggarkan prestasi pelajar di UTHM dan institusi pendidikan tinggi yang seumpamanya.

## TABLE OF CONTENTS

<b>TITLE</b>	<b>i</b>
<b>DECLARATION</b>	<b>ii</b>
<b>DEDICATION</b>	<b>iii</b>
<b>ACKNOWLEDGEMENT</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>ABSTRAK</b>	<b>vi</b>
<b>TABLE OF CONTENTS</b>	<b>vii</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>LIST OF SYMBOLS AND ABBREVIATIONS</b>	<b>xii</b>
<b>LIST OF APPENDICES</b>	<b>xv</b>
<b>LIST OF PUBLICATIONS</b>	<b>xvi</b>
<b>CHAPTER 1 INTRODUCTION</b>	<b>1</b>
1.1 Research background	1
1.2 Problem statement	3
1.3 Research objectives	5
1.4 Research scope	5
1.5 Research significance	6
1.6 Thesis organization	6
<b>CHAPTER 2 LITERATURE REVIEW</b>	<b>7</b>
2.1 Students performance	7
2.2 Statistical techniques applied to study on student performance	9
2.3 Survival analysis for student performance	10
2.4 Survival analysis in other sector	12
2.5 Research gap	14

<b>CHAPTER 3</b>	<b>METHODOLOGY</b>	<b>20</b>
3.1	Data collection	20
3.2	Data description	21
3.3	Kaplan-Meier (KM) method	23
3.3.1	Log-rank test	24
3.3.2	Wilcoxon test	24
3.3.3	Tarone-Ware test	25
3.4	Cox Model	26
3.4.1	Hazard ratio	27
3.4.2	Cox-Snell residual	28
3.4.3	Schoenfeld residual	28
3.4.4	Martingale residual	29
3.5	Parametric Model	30
3.5.1	Parametric survival model	30
3.5.1.1	Exponential model	30
3.5.1.2	Weibull model	31
3.5.1.3	Log-logistic model	32
3.5.1.4	Log-normal model	33
3.5.2	Parametric AFT Model	34
3.5.2.1	The general AFT model	34
3.5.2.2	Log-linear form AFT model	35
3.5.2.3	Exponential AFT model	37
3.5.2.4	Weibull AFT model	38
3.5.2.5	Log-logistic AFT model	39
3.5.2.6	Log-normal AFT model	41
3.5.2.7	Time ratio (TR)	43
3.6	Model Selection	43
3.6.1	Log-likelihood value	43
3.6.2	Akaike Information Criteria (AIC)	44
3.6.3	Bayesian Information Criteria (BIC)	44
3.6.4	Corrected Akaike Information Criteria	45
3.7	Research Framework	45

<b>CHAPTER 4</b>	<b>RESULT AND DISCUSSION</b>	<b>48</b>
4.1	Descriptive analysis	48
4.2	Nonparametric survival analysis	49
4.2.1	KM analysis	49
4.2.2	Comparison of survival result	54
4.3	Cox model analysis	55
4.3.1	Covariates selection by Cox model	55
4.3.2	Hazard ratio for Cox model	56
4.3.3	Cox model residual analysis	57
4.4	Parametric model	63
4.4.1	Parametric survival model	63
4.4.2	Parametric AFT model	66
4.4.3	Comparative result of the parametric survival and parametric AFT model	66
4.4.4	The analysis of covariates of the best model	67
<b>CHAPTER 5</b>	<b>CONCLUSION</b>	<b>69</b>
5.1	Main findings	69
5.2	Research contribution	71
5.3	Limitation and recommendation	72
	<b>REFERENCES</b>	<b>73</b>
	<b>APPENDICES</b>	<b>86</b>
	<b>VITA</b>	



## LIST OF TABLES

2.1	The summary of literature review	15
3.1	Description of variables and codes for data 2011/2012	21
3.2	Data of Bachelor's degree student	22
3.3	Student-semester data set	22
4.1	Number of students based on entrance qualification for each semester	49
4.2	Probability of survival and failure for diploma-qualification	51
4.3	Probability of survival and failure for matriculation-qualification	52
4.4	Probability of survival and failure for STPM-qualification	52
4.5	$\chi^2$ test and $p$ -value for three entrances qualification	54
4.6	AIC and $p$ -value for covariates selection in Cox PH model	55
4.7	Analysis of the covariates that contribute to student performance	57
4.8	Scaled Schoenfeld residual of covariates in the Cox model	60
4.9	Log-likelihood, AIC, AIC <sub>C</sub> and BIC values of parametric survival models	65
4.10	Log-likelihood, AIC, AIC <sub>C</sub> and BIC values of parametric AFT models	66
4.11	Comparison results of best parametric and parametric AFT model	67
4.12	Coefficient estimates, TR and $p$ -value for log-normal AFT model	67

## LIST OF FIGURES

3.1	Research framework on student performance in UTHM	47
4.1	Survival probability plot by using KM curve	50
4.2	Survival and failure probability plot for diploma, matriculation and STPM	53
4.3	Cumulative hazards plot of Cox-Snell residuals	58
4.4(a)	Scaled Schoenfeld residuals plot for GPA	59
4.4(b)	Scaled Schoenfeld residuals plot for entrance qualification	59
4.4(c)	Scaled Schoenfeld residuals plot for course	60
4.5(a)	Plot of covariates GPA against Martingale residual	61
4.5(b)	Plot of covariates qualifications against Martingale residual	62
4.5(c)	Plot of covariates courses against Martingale residual	62
4.6(a)	Survival probability plot for exponential fitted with KM plot	63
4.6(b)	Survival probability plot for Weibull fitted with KM plot	64
4.6(c)	Survival probability plot for log-normal fitted with KM plot	64
4.6(d)	Survival probability plot for log-logistic fitted with KM plot	65

## LIST OF SYMBOLS AND ABBREVIATIONS

$a_i$	-	Coefficients estimate of the covariates
$e_{ij}$	-	Expression of proportion groups at risk to fail by number of failure
$q - 1$	-	Degrees of freedom
$i$	-	Groups or covariates in the study, $i = 1, 2, \dots, n$
$j$	-	Failure time, $j = 1, 2, \dots, p$
$k$	-	Number of constant parameter of each parametric model
$m_i$	-	Number of failure that happened at time $t_i$
$m_{ij}$	-	Number of failure that happened by groups, $i$ at $j^{th}$ failure time
$n$	-	Number of observations in the data set
$n_i$	-	Number of observations at risk at the beginning of time period $t_i$
$n_{ij}$	-	Number of observations at risk by groups, $i$ at $j^{th}$ failure time
$q$	-	Number of covariates in the model
$t$	-	Time of study
$t_i$	-	A time when an event happened by group in a study
$t_j$	-	A time when failure happened at $j^{th}$ failure time
$T_i$	-	Log-linear model for the random variable associated with the lifetime of the $i^{th}$ observation in a survival study
$x_i$	-	Covariate values in the study by groups
$X_{ij}$	-	Covariate values in the study by groups at $j^{th}$ failure time
$Y_i$	-	Process determining if the $i^{th}$ subject at risk to fail at time, $t$
$f(t)$	-	Probability density function
$F(t)$	-	Cumulative distribution functions
$h(t)$	-	Hazard function
$h_0(t)$	-	Baseline hazard function

$H_0$	-	Cumulative baseline hazard function
$\hat{H}_0$	-	Estimate values of cumulative baseline hazard function
$S(t)$	-	Survival function
$S_0(t)$	-	Baseline survival function
$w(t_j)$	-	Weight at a time when failure happened at $j^{th}$ failure time
$\lambda$	-	Scaling parameter
$\gamma$	-	Shape parameter for parametric model
$\theta$	-	Location parameter
$\mu$	-	Mean log of distribution, intercept parameter for parametric AFT model
$\sigma$	-	Shape parameter for parametric AFT model
$\Phi$	-	Cumulative distribution function of the normal distribution
$\beta$	-	Residual value
$\hat{\beta}$	-	Estimated residual value
$\delta_i$	-	Define status for the event
$\epsilon_i$	-	Random variable, assumed to have a particular probability distribution
$\eta_i$	-	Linear component of the model
AMO	-	Academic Management Office
AFT	-	Accelerated Failure Time
AIC	-	Akaike Information Criterion
AIC <sub>C</sub>	-	Corrected Akaike Information Criterion
BIC	-	Bayesian Information Criterion
CeDS	-	Centre for Diploma Studies (CeDs)
CGS	-	Centre of Graduate Studies (CGS)
FKAAS	-	Faculty of Civil and Environmental Engineering
FKEE	-	Faculty of Electrical and Electronic Engineering
FKMP	-	Faculty of Mechanical and Manufacturing Engineering
FPTP	-	Faculty of Technology Management and Business
FPTV	-	Faculty of Technical and Vocational Education
FSKTM	-	Faculty of Computer Science and Information Technology
GPA	-	Grade Point Average
HR	-	Hazard ratio
KM	-	Kaplan Meier
PH	-	Proportional Hazard

PPA	-	Pejabat Pengurusan Akademik
SPSS	-	Statistical Package for the Social Sciences
STPM	-	Malaysian Higher School Certificate
TR	-	Time ratio
UTHM	-	Universiti Tun Hussein Onn Malaysia



## LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Call the variables in R	86
B	Kaplan Meier analysis by using R	87
C	Non-parametric statistical test	89
D	Covariates selection for Cox model	91
E	Fit Cox model and estimate hazard ratio	94
F	Testing the Cox Proportional Hazards assumptions	95
G	Parametric model and Parametric AFT model	97
H	Estimate time ratio for the best model	101



PT TAAUTHM  
PERPUSTAKAAN TUNKU TUN AMINAH

## LIST OF PUBLICATIONS

- Azme Khamis, Che Azmeeza Che Hamat, Mohd Asrul Affendi Abdullah, Sabariah Saharan, Norhaidah Mohd Asrah. (2017). *International Journal of Advanced Research in Science, Engineering and Technology*. 4649-4654.
- Azme Khamis & Che Azmeeza Che Hamat. (2018). Non-Parametric Test for Detection Trend in Crude Palm Oil Price. *Scientific Research Journal*, 11-18.
- Azme Khamis, Che Azmeeza Che Hamat, Mohd Asrul Affendi Abdullah. (2018). The Post Hoc Procedure in Survival Analysis for Undergraduate Students Performance. *Journal of Science and Technology*, 175-183.



PTTA  
PERPUSTAKAAN TUNKU TUN AMINAH

## CHAPTER 1

### INTRODUCTION

This chapter introduces the research background that mainly focuses on student performance. Overviews of the survival analysis application along with the significant models that are relevant to the present study are discussed to provide insight of the main critical topic of the study. The problem statement, research objective and research significance are also presented in this chapter.

#### 1.1 Research background

Student performance is an important issue in education especially in higher learning institutions. Kyoshiba (2009) defined student performance as students who are capable to succeed during their studies. Student performance is the outcome of education when students achieved their educational goals. Performance is characterised by performance in a case method course which may assess along various dimensions including class participation, student work on papers and exams, and group activities such as projects and presentations (Steinmayr *et al.*, 2014). Mat *et al.* (2014) characterised student performance by exam oriented performance and represented by grade point average (GPA) once the students complete the study.

Performance of students can be measured by estimating the covariates of characteristics that affect student performance. In general, covariates are characteristics, factors or explanatory variables that exist within the study. Investigations of covariates on the performance of students have become a topic of growing interest in higher educational circle. Recent studies were carried out to



explore the covariates that affected university student performance. These covariates may range from student education background (Boghikian-Whitby & Mortagy, 2008), previous school prior to admission into a programme (Kyoshaba, 2009), GPA (Bruinsma & Jansen, 2009) and various entry qualifications obtained by the students for admission into a higher academic programme (Amasuomo, 2014). The other covariates that contribute to student performance are geographical background (Hovdhaugen, 2011), gender (Zuilkowski & Jukes, 2014) and study habits (Shahzadi & Ahmad, 2011).

Study on student performance can be conducted using various statistical techniques including regression analysis (Hijazi & Naqvi, 2006), categorical regression analysis (Cilan & Can, 2014), data mining techniques of the hierarchical linear modelling (Miller *et al.*, 2015), and meta-analysis approach (Chang, Jian & Chou, 2016). These studies were carried out to identify the most significant factors or covariates affecting the student performance and identify any risk of failure among students during their study. Survival techniques have applied in education to focus on student failure rate in higher education study program (Paura & Arhipova, 2014). The study found that survival analysis technique showed excellent estimation of student performance by applying time together with discrete or continuous independent variables. Survival analysis also has binary outcome variables that can be used to detect the status of students during the study.

In survival analysis, Kaplan-Meier (KM) method is a simple method (Goel, Khanna & Kishore, 2010) as it has no ability to test or investigate the covariates effect on student performance. KM method is limited to analyse one covariate in each study. However, existence of many covariates in a study may affect the outcome of the study (Clark *et al.*, 2003). The survival analysis proposes the use of the Cox model to estimate the hazard ratio (HR) of each covariate. The estimation in Cox model is biased and the results derived from the model may be erroneous when the proportional hazard (PH) assumption in the Cox model is violated (Persson & Khamis, 2005). The biased estimation is likely because a PH assumption in Cox model is the HR for each covariate and it is always assumed as constant overtime.

Violation of PH assumption causes the unreliable and biased results from the Cox model. Therefore, Accelerated Failure Time (AFT) models such as exponential, Weibull, log-logistic and log-normal are better choices to apply in the study. As AFT models consider a statistical distribution for survival time and do not rely on PH

assumptions, they are deemed as suitable alternatives to Cox model (Zare *et al.*, 2015). In order to overcome the weaknesses in Cox model, the parametric AFT models is applied in the study to estimate the effect of covariates on accelerating or decelerating the student survival or known as time ratio (TR).

The aims of this study are to propose the best survival model to model and formulate a good conclusion to the covariates that exist in the study of student performance. Besides, this study also estimates the HR and TR of covariates and determines the effect of the covariates to student performance. The novelty of this research focuses on suggesting various analytical techniques on student performance by using survival techniques such as KM method, Cox model and parametric AFT model.

## 1.2 Problem statement

Study on student performance comes in various techniques such as regression analysis, categorical regression analysis, data mining techniques and meta-analysis approach. The techniques that are applied in previous studies of student performance are limited to only one categorical variable and numeric variable for covariate (Cilan & Can, 2014). The missing data are ignored in the calculation of student performance (Miller *et al.*, 2015). Limitations from previous studies of student performance prompted the study to use the survival techniques. Survival technique is the simplest method to find the survival rate or student performance without ignorant the missing data.

The benefit of the non-parametric via KM method in survival techniques lies in its flexibility to include information about students who are lost at any point in time which can be used to estimate the survival rates (Leonavicius, 2009). Various studies have been explored on the estimate of student performance according to the demographic characteristic (Zuilkowski & Jukes, 2014) and geographical background (Hovdhaugen, 2011). However, within the context of the effect of covariates proposed in the study of student performance in UTHM, very little studies in the literature applied have been dedicated to study the effect of entrance qualification as an important covariate and estimate survival rate in the research based on the covariate.

Estimate survival rate in the study of student performance using KM method was not able to estimate the multiple covariates used in the study at a time of analysis.

By using semi-parametric model via Cox model, the effect of covariates on the performance of student can be investigated. Thus, many covariates can be utilised to seek information within the sets of data pertaining the student at risk to fail or calculation of students HR. Previous study has been considered type of school, gender, age and GPA covariates together in the study of student performance (Bruinsma & Jansen, 2009). Gender, grades and course covariates have employed by Min *et al.* (2011) while Box-Steffensmeir *et al.* (2015) have considered faculty and gender covariates in the study of student performance. No studies in the literature have combined the faculty, course and student grades covariates in the study of student performance. Incorporation of these variables prompts the question on which covariates dominate the performance of students and can be utilised for further analysis.

Estimate of student at risk to fail or a HR of students through the existing covariates demonstrates the expectations about the future developments of student performance. A related question is how the Cox model can provide an accurate result for this estimated based for the covariates. By using semi-parametric model via PH assumption through the residual analysis, the performance of Cox model in analysing student performance can be investigated. When the PH assumption in Cox model is violated, the result provided by Cox model is considered as less accurate (Baghestani *et al.*, 2015).

The parametric models have many distributions that can be utilised in the survival analysis. Very few studies in the literature have applied the parametric survival model and parametric AFT model in accessing student performance. Hence, it is questionable which of the models gives a good fitted data and can be used for further analysis. The best model will be selected to further study by using model selection method to select the parametric and the parametric AFT models.

Estimate of students' ability to survive or TR of students through significant covariates will increase the validation of the study. Henceforth, the study will need to ascertain which the best model can estimate the TR of each significant covariate to conclude the overall study. The TR on student performance can be estimated by using the best model. Therefore, the aim of this study is to propose the best alternative model in evaluating the student performance.

### 1.3 Research objective

This study aim to achieve the following objectives:

1. To estimate the survival rates of students with different entrance qualifications using KM method.
2. To identify the covariates that dominate the student performance via Cox model.
3. To investigate the performance of Cox model based on the violation of the PH assumptions.
4. To compare the model performance by using the parametric survival and parametric AFT models.
5. To estimate the time ratio (TR) of student performance in accordance to the best model selected in Objective 4.

### 1.4 Research scope

This research involves Bachelor's degree students in UTHM. Total numbers of Bachelor's degree students in UTHM of 2011/2012 intake are 2606 based on standard entrance qualifications that are diploma, matriculation and STPM. All 2606 Bachelor's degree students for 2011/2012 intake are collected from the Academic Management Office (AMO) in UTHM and analysed by using Statistical Package for the Social Sciences (SPSS) and R statistical software. In 2011, UTHM only have six faculties for Bachelor's degree students namely the Faculty of Technical and Vocational Education (FPTV), Faculty of Electrical and Electronic Engineering (FKEE), Faculty of Civil and Environmental Engineering (FKAAS), Faculty of Mechanical and Manufacturing Engineering (FKMP), Faculty of Computer Science and Information Technology (FSKTM) and Faculty of Technology Management and Business (FPTP). All Bachelor's Degree students from these six faculties for 2011/2012 intake are involved in the study of student performance. Centre for Diploma Studies (CeDS) and Centre of Graduate Studies (CGS) are excluded since this study focused only on Bachelor's Degree students.

This study focuses on the potential of student failure in each semester. The event of interest in this study is student failure with GPA of less than 2.00 for the current semester of study. Students are considered as censored if they achieved any GPA value

that is greater or equal to 2.00. The dependent variable is the time of the event of interest occurs. It is defined as 1 for the event status with GPA less than 2.00 and 0 if it is censored with GPA greater and equal to 2.00. The independent variables are the entrance qualification, faculty, course and the GPA of students.

This study considered several survival approaches including non-parametric, semi-parametric and parametric, for example KM method, Cox model, exponential model, Weibull model, log-logistic model, log-normal model, exponential AFT model, Weibull AFT model, log-logistic AFT model and log-normal AFT model.

### **1.5 Research significance**

This research aims to estimate the survival rate of students based on different entrance qualification. The Cox model applied design to identify the covariates that dominate the student performance. This study will investigate the performance of the Cox model based on PH assumption, whether the PH assumption in the Cox model is accepted or violated. The parametric model involved is intended to overcome the limitations encountered by the Cox model. The parametric AFT model is the extension method of the parametric survival model that is capable to estimate the TR for significant covariates in the study. Finally, this study plans to compare the student performance by using parametric survival and parametric AFT model in a way to propose an alternative survival model to evaluate the student performance.

### **1.6 Thesis organization**

Chapter 2 which provides a literature review on student performance and survival analysis. Chapter 3 presents the data set and its description where the research methodologies that explain well before performed the survival analysis. Chapter 4 summarises and discusses the overall results. Chapter 5 discusses and concludes the main finding of the study, highlights the research contribution and limitations, and provides recommendations for future study on the similar topic.

## **CHAPTER 2**

### **LITERATURE REVIEW**

This chapter discusses the previous studies on student performance and the methods that have been used in the survival analysis. It also provides the reader with the survival techniques that have been applied to the higher education and other sectors. The first part of this chapter gives a brief description of the past and current studies conducted on student performance followed by a brief review of various statistical methods. Finally, the research gap on the application of statistical techniques in education is discussed in this chapter.

#### **2.1 Student performance**

Hijazi & Naqvi (2006) carried out a case study on factors affecting student performance and found that study hour was not a significant factor as far as student performance is concerned. Student performance may depend on intelligence level, intellect, memory or learning method of the students. Meanwhile, Rasul & Bukhsh (2011) examined a relationship among psychological, physical, socio-economic and educational factors on the performance of student in examination. The study found that most of the psychological, physical, socio-economic and educational factors affected the performance of students in examination at university level.

Higher education institutions have always greatly emphasised on the factors governing the performance of students. Lesik (2007) performed a study to identify a casual impact of education programs on student performance. He stated that most of the existing research has been non-experimental and focused exclusively on student background and demographic characteristics to study student success. Belanger & Liu



(2008) also agreed that household characterised with high income impacted student performance. Cilan & Can (2014) conducted a study to measure the factors affecting student academic performance and found that demographic factors, which are age and student grades, are the factors that significantly affected students' academic success. These studies offered a framework that student background, demographic characteristics, and student grades have causal impact on student success.

Shahzadi & Ahmad (2011) found that academic interaction with teachers and peers had significant effect on the student academic performance. Home environment and study habits were also important factors in the student academic performance at university level. Students must improve their study habits especially allocating proper study time every day and revise the lecture notes on daily basis. The research concluded that better study habits prompted higher academic interaction and ultimately led to higher academic performance of students. Chang *et al.* (2016) conducted a study to identify the effects of student engagement on academic achievement and revealed strong relationship existing between student behavioural, student engagement and academic achievement.

The academic achievements determined the factors that contributed to student performance in the university. Tianqi & Ganges (1995) conducted a study to examine the performance of students. The study included student gender, ethnicity, graduation date and GPA as the determinants and found that students' gender and ethnicity affected on student performance. The second, third and fourth semesters are always that most critical periods for the students to leave their study. In 1999, Murtaugh, Burns & Schuster (1999) conducted a study on undergraduate students by considering the students' college residency, ethnicity, age and GPA covariates to examine the performance of the students. Age, high GPA and non-residents college students were shown to have higher performance rates than the resident and international students. This analysis helped the university to guide and improve student performance in the university.

Leonavicius (2009) also conducted a study focusing on student performance and student retention. This study used gender and entrance grade of each student as the covariates. The study found that gender is not a significant factor for study on student performance and student retention. However, Bruinsma & Jansen (2009) revealed that 61% of the students in their study passed their first year examination within 24 months, with female students passing it earlier than male students and those

with higher GPA passed their examination earlier than students with lower GPA. Min *et al.* (2011) focused on the survival analysis of the loss rate of undergraduate students. This study included grades, genders, course of each student, and found the existence of statistical differences between levels of grades received by different courses. Additionally, female students tend to leave engineering course study earlier than their male counterparts.

Paura & Arhipova (2014) analysed the student performance rate in higher education study program by using genders, grades and financial source covariates. The results of this study showed that the main reasons for student non-perform were student low grades during the study and low motivation to continue the study. In 2015, a research on student performance was conducted at the university level by using the year of study (Kartal & Ozan, 2015) and faculty (Box-Steffensmeier *et al.*, 2015) covariates. The study showed that performance of students neither depended on the year of study nor the faculty that the students belonged to.

## **2.2 Statistical techniques applied to study on student performance**

Various methods have been explored regarding the study on student performance such as regression analysis (Hijazi & Naqvi, 2006), mean score (Rasul & Bukhsh, 2011), confirmatory factor analysis (Shahzadi & Ahmad, 2011), categorical regression analysis (Cilan & Can, 2014) and meta-analysis approach (Chang *et al.*, 2016).

Regression analysis was used to explore the factors that affected student performance. Hijazi & Naqvi (2006) applied regression analysis for the study on student performance, dealing with student attitude towards study, attendance, home environment and previous academic achievements as the covariates. The study concluded that the student attitude towards study was positively related to student performance.

Rasul & Bukhsh (2011) applied the mean score to determine the performance of students in examination considering the examination is an integral part of the education system. Examination is useful as it measures a student progress and acts as a process to test the abilities or achievement of the student in any area of academic programs. Examination system has many factors that directly and indirectly affect student performance. Shahzadi & Ahmad (2011) applied the confirmatory factor analysis and developed structural equation models to access the student academic



## REFERENCES

- Aalen, O.O. & Johansen, S. (1978). An Empirical Transition Matrix for the Nonhomogeneous the Markov Chains Based on the Censored Observations. *Scandinavian Journal of Statistics*, 5, pp. 141-150.
- Aalen, O.O., Roysland, K., Gran, J.M. & Ledergerber, B. (2012). Causality, Mediation and Time: A Dynamic Viewpoint. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 175(8), pp. 31–61.
- Abner, E.L., Charnigo, R.J. & Kryscio, R.J. (2014). Markov Chains and Semi-Markov Models in Time-to-Event Analysis. *Journal of Biometrics and Biostatistics*, 1(19522), pp. 1-5.
- Abrahamowicz, M., Mackenzie, T. & Esdaile, J.M. (1996). Time-dependent Hazard Ratio: Modeling and Hypothesis Testing with Application in Lupus Nephritis. *Journal of American Statistical Association*, 91(433), pp. 1432–1439.
- Akaike, H.A.I. (1974). A New Look at the Statistical Model Identification. *IEEE Transaction on Automatic Control*, 19(6), pp. 716-723.
- Alarcon, G.M. & Edwards, J.M. (2013). Ability and Motivation: Assessing Individual Factors That Contribute to University Retention. *Journal of Educational Psychology*, 105(1), pp. 129-137.
- Allison, P. (1984). *Event History Analysis: Regression and Longitudinal Data*. Beverly Hills: Sage Publications.
- Alhasawi, E. (2015). *Survival Analysis Approaches for Prostate Cancer*. Laurentian University. Master's Thesis.
- Al-Shomrani, A.A., Shawky, A.I., Arif, O.H. & Aslam, M. (2016). Log-logistic Distribution for Survival Data Analysis using Markov Chain Monte Carlo (MCMC) Technique. *Springer Plus*, 5(1774), pp. 1-16.

- Amasuomo, J.O.M. (2014). Academic Performance of Students Admitted with Different Entry Certificates to The Nigerian Certificate in Education Programme at The Federal College of Education (Technical), Omoku. *Journal of Technology and Science Education*, 4(1), pp. 39-47.
- Ameri, S., Fard, J.M., Chinnam, R.B. & Reddy, C.K. (2016). Survival Analysis Based Framework for Early Prediction of Students Dropout. *Proceeding of the 25<sup>th</sup> ACM International on Conference on Information and Knowledge Management*. pp. 903-912.
- Andersen, P.K. & Gill, R.D. (1982). Cox Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4), pp. 1100-1120.
- Ansin, E. (2015). *An Evaluation of the Cox-Snell Residuals*. Uppsala University: Master's Thesis.
- Baghestani, A.R., Moghaddam, S.S., Majd, H.A., Akbari, M.E., Nafissi, N. & Gohari, K. (2015). Survival Analysis of Patients with Breast Cancer using Weibull Parametric Model. *Asian Pacific Journal of Cancer Prevention*, 16(18), pp. 8567-8571.
- Barlow, W.E. & Prentice, R.L. (1988). Residuals for Relative Risk Regression. *Biometrika*, 75(1), pp 65-74.
- Bedrick, E.J. & Tsai, C.L. (1994). Model Selection for Multivariate Regression in Small Samples. *Biometrics*, 50(1), pp. 226-231.
- Belanger, D. & Liu, J. (2008). Education and Inequalities in Rural Vietnam in the 1990s. *Asia Pacific Journal of Education*, 28(1), pp. 51-65.
- Bewick, V., Cheek, L. & Ball, J. (2004). Statistics review 12: Survival analysis. *BioMed Central: Critical Care*, 8(5), pp. 389-394.
- Bland, J.M. & Altman, D.G. (1998). Survival probabilities (The Kaplan-Meier Method). *British Medical Journal*, 317(7172), pp. 1572-1580.
- Boghikian-Whitby, S. & Mortagy, Y. (2008). The Effect of Student Background in E-Learning-Longitudinal Study. *Issues in Informing Science and Information Technology*, 5, pp. 107-126.
- Borges, G., Medina, M.I.M.E., Benjet, C., Lee, S., Lane, M. & Breslau, J. (2011). Influence of Mental Disorders on School Dropout in Mexico. *Pan American Journal of Public Health*, 30(5), pp. 477-483.

- Bowers, A.J. (2010). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *The Journal of Educational Research*, 103(3), pp. 191-207.
- Box-Steffensmeier, J.M., Cunha, R.C., Varbanov, R.A., Hoh, Y.S., Knisley, M.L. & Holmes, M.A. (2015). Survival Analysis of Faculty Retention and Promotion in the Social Sciences by Gender. *Public Library of Science (PLOS) Journal*, 10(11), doi: 10.1371/journal.pone.0143093.
- Bradburn, M.J., Clark, T.G., Love, S.B. & Altman, D.G. (2003). Survival Analysis Part II: Multivariate Data Analysis – An Introduction to Concepts and Methods. *British Journal of Cancer*, 89(3), pp. 431 – 436.
- Breslow, N. (1970). A generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. *Biometrika*, 57(3), pp. 579-594.
- Breslow, N. & Crowley, J. (1974). A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship. *The Annals of Statistics*, 2(3), pp. 437-453.
- Brezavscek, A., Bach, M.P. & Baggia, A. (2017). Markov Analysis of Students Performance and Academic Progress in higher Education. *Organizacija*, 50(2), pp. 83-96.
- Bruinsma, M. & Jansen, E.P.W.A. (2009). When will I Succeed in My First-Year Diploma? Survival Analysis in Dutch Higher Education. *Higher Education Research & Development*, 28(1), pp. 99-114.
- Cavanaugh, J.E. (1997). Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria. *Statistics & Probability Letters*, 33(2), pp. 201-208.
- Chang, D.F., Jian, W.C. & Chou, W.C. (2016). Meta-analysis Approach to Detect the Effect of Student Engagement on Academic Achievement. *11<sup>th</sup> International Conference on Innovative Computing, Information and Control 2016 Proceeding*. Tamkang University Institutional Repository. pp. 1-6.
- Chatfield, C. (1995). Model Uncertainty, Data Mining and Statistical Inference. *Journal of Royal Statistical Society*, 158(3), pp. 419-466.
- Chimka, R. & Lauren, H. (2008). Interaction and Survival Analysis of Graduation Data. *Academic Journals*, 3(1), pp. 29-32.
- Christopher, H.J. (2016). Flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, 70(8), pp. 1-33.

- Cilan, C.A. & Can, M. (2014). Measuring factors Effecting MBA Students Academic Performance by Using Categorical Regression Analysis: A Case Study of Institution of Business Economics, Istanbul University. *Procedia Social and Behavioral Sciences*, 122(2014), pp. 405-409.
- Clark, T.G., Bradburn, M.J., Love, S.B. & Altman, D.G. (2003). Survival Analysis Part IV: Further Concepts and Methods in Survival Analysis. *British Journal of Cancer*, 89(5), pp. 781-786.
- Cohen, A.C. (1965). Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples. *Journal of Technometrics*, 7(4), pp. 579-588.
- Collett, D. (1994.) *Modelling Survival Data in Medical Research*. 1<sup>st</sup> ed. London: Chapman and Hall.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. 2<sup>nd</sup> ed. London: Chapman and Hall.
- Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, 34(2), pp.187-220.
- Cox, D.R. & Oakes, D. (1984). *Analysis of survival data*. London: Chapman and Hall.
- Creswell, J.W. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 2<sup>nd</sup> ed. Los Angeles: Sage Publications.
- Darden, W.R., Hampton, R.D. & Boatwright, E.W. (1987). An Application of Survival Analysis. *Journal of Retailing*, 63(1), pp. 69-88.
- Davidian M. & Louis T. (2012). *Why Statistics?* Washington: Science.
- Elisa, T.L. & Wang, J.W. (1992). *Statistical Methods for Survival Data Analysis*. 2<sup>nd</sup> ed. United Kingdom: A Wiley-Interscience Publication.
- Etzioni, R.D., Feuer, E.J., Sullivan, S.D., Lin, D., Hu, C. & Ramsey, S.D. (1999). On the use of Survival Analysis Techniques to Estimate Medical Care Cost. *Journal of Health Economics*, 18(3), pp. 365-380.
- Fan, J. & Li, R. (2002). Variable Selection for Cox's Proportional Hazards Model and Frailty Model. *The Annals of Statistics*, 30(1), pp. 74-99.
- Faraggi, D. & Simon, R. (1998). Bayesian Variable Selection Method for Censored Survival Data. *Biometrics*, 54(4), pp. 1475-1485.
- Farooq, M.S., Chaudhry, A.H., Shafiq, M. & Berhanu, G. (2011). Factor Affecting Students' Quality of Academic Performance: A Case of Secondary School Level. *Journal of Quality and Technology Management*, 7(2), pp. 1-14.

- Feinleib, M.A. (1960). A Method of Analysing Log-normally Distributed Survival Data with Incomplete Follow-Up. *Journal of the American Statistical Association*, 55(291), pp. 534-545.
- Fisher, L.D. & Lin, D.Y. (1999). Time-dependent Covariates in The Cox Proportional Hazards Regression Model. *Annual Review of Public Health*, 20(1), pp. 145-157.
- Fitrianto, A. (2013). Several Types of Residuals in Cox Regression Model: An Empirical Study. *International Journal of Mathematics Analysis*, 7(53). pp. 2645-2654.
- Frank, E. & Harrel, Jr. (2001). Regression Modelling Strategies: With applications to Linear Models, Logistic Regression and Survival Analysis. *Journal of the American Statistical Association*, 98(1), pp. 257-258.
- Gabriel, D., Brigitte, S.C. & Elisabeta, J. (2017). Estimation of Unemployment Duration in Botosani County using Survival Analysis. *Economic Sciences Series*, 17(1), pp. 155-161.
- Garcia, R.I., Ibrahim, J.G. & Zhu, H. (2010). Variable Selection in the Cox Regression Model with Covariates Missing at Random. *Biometrics*, 66(1), pp. 97-104.
- Gehan, E.A. (1965). A Generalized Wilcoxon Test for Comparing Arbitrarily Singly Censored Samples. *Biometrika*, 52(39), pp. 203-213.
- George B., Seals, S & Aban, I. (2014). Survival Analysis and Regression Models. *Journal of Nuclear Cardiology*, 21(4), pp. 686–694.
- Gerdes, H. & Mallinckrodt, B. (1994). Emotional, Social, and Academic Adjustment of College Students: A Longitudinal Study of Retention. *Journal of Counseling and Development*, 72(3), pp. 281-288.
- Goel, M.K., Khanna, P. & Kishore, J. (2010). Understanding Survival Analysis: Kaplan-Meier Estimate. *International Journal of Ayuverda Reseacrh*, 1(4), pp. 274–278.
- Greenhouse, J., Stangl, D. & Bromberg, J. (1989). An Introduction to Survival Analysis: Statistics Method for Analysis of Clinical Trial Data. *Journal of consulting clinical psychology*, 57(4), pp. 536-544.
- Harrell, F.E. (2001). *Regression Modelling Strategies, with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer.

- Hess, K.R. (1995). Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression. *Statistics in Medicine*, 14(15), pp. 1707-1723.
- Hijazi, S.T. & Naqvi, S.M.M.R. (2006). Factors Affecting Students' Performance A Case of Private Colleges. *Bangladesh e-Journal of Sociology*, 3(1), pp. 1-10.
- Hosmer, D.W. & Lemeshow, S. (1999). *Applied Survival Analysis*. 1<sup>st</sup> ed. New York: John Wiley and Sons.
- Hovdhaugen, E. (2011). Do Structured Study Programmers Lead to Lower Rates of Dropout and Student Transfer from University? *Irish Educational Studies*, 30(2), pp. 237-251.
- Hoverstad, R., Sylvester, R. & Voss, K.E. (2001). The Expected Monetary Value of a Student, a Model and Example. *Journal of Marketing for Higher Education*, 10(4), pp. 51-62.
- Hurvich, C.M. & Tsai, C.L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2), pp. 297-307.
- Hurvich, C.M., Shumway, R.H. & Tsai, C.L. (1990). Improved Estimators of Kullback-Leibler Information for Autoregressive Model Selection in Small Samples, *Biometrika*, 77(4), pp. 709-719.
- Hutton, J.L. & Monaghan, P.F. (2002). Choice of Parametric Accelerated Life and Proportional Hazards Models for Survival Data: Asymptotic Results. *Lifetime Data Analysis*, 8(4), pp. 375-393.
- Jadwiga, B. (2013). Extensions of Cox Model for Non-Proportional Hazards Purpose. *PhUSE*, <https://www.lexjansen.com/phuse/2013/sp/SP07.pdf>.
- Kalbfleisch, J.D. & Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. 2<sup>nd</sup> ed. New York: John Wiley & Sons.
- Kaplan, E.L. & Meier, P. (1958). Non-parametric Estimation from Incomplete Observations. *Journal of American Statistical Association*, 53(282), pp. 457-481.
- Kartal, M. & Ozan, M.S. (2015). *Using Survival Analysis to Investigate the Persistence of Students in an Introductory Information Technology Course at Metu*. Middle East Technical University: Master's Thesis.



- Kay, R. & Kinnersley, N. (2002). On the use of Accelerated Failure Time Model as an Alternative to the Proportional Hazards Model in the Treatment of Time to Event Data: A Case Study in Influenza. *Drug Information Journal*, 36, pp. 571-579.
- Klein, J.P. (1991). Small sample moments of some estimators of the variance of the Kaplan-Meier and Nelson Aalen estimators. *Scandinavian Journal of Statistics*, 18(4), pp. 333-340.
- Klein, J.P. & Moeschberger, M.L. (1997). *Survival Analysis, Technique for Censored and Truncated Data*. 1<sup>st</sup> ed. New York: Springer.
- Klein, J.P. & Moeschberger M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2<sup>nd</sup> ed. United States: Springer.
- Kleinbaum, D.G. (1996). *Survival Analysis: A Self-Learning Text*. New York: Springer.
- Kleinbaum, D.G. & Klein, M. (2005). *Survival Analysis a Self-Learning Text*. New York: Springer.
- Kyoshaba, M. (2009). *Factors Affecting Academic Performance of Undergraduate Students at Uganda Christian University*. University of Makerere. Master's Thesis.
- Laungerman, M., Shelley, M.C., Diane, R. & Mickelson, S.K. (2015). Estimating Survival Rates in Engineering for Community College Transfer Students using Grades in Calculus and Physics. *International Journal of Education in Mathematics, Science, and Technology*, 3(4), pp. 313-321.
- Leonaviciene, T. (2009). Analysis of the Study Process using Survival Analysis Methods. *Pedagogika Studies*, 93, pp. 98-110.
- Leonavicius, G. (2009). Research into Bachelor's Degree Studies of Informatics at VPU using Survival Analysis Methods. *Pedagogika Studies*, 94, pp. 95-98.
- Lesik, S.A. (2007). Do Developmental Mathematics Programs Have a Causal Impact on Student Retention? An Application of Discrete-Time Survival and Regression-Discontinuity Analysis. *Research in Higher Education*, 48(5), pp. 583-607.
- Li, S. (1996). Survival Analysis. *Marketing Research*, 7(4), pp. 17-23.
- Liang, H. & Zou, G. (2008). Improved AIC Selection Strategy for Survival Analysis. *National Institutes of Health (NIH) Public Access*, 52(5), pp. 2538-2548.

- Lin, D.Y., Wei, L.J. & Ying, Z. (1998). Accelerated Failure Time Models for Counting Processes. *Biometrika*, 85(3), pp. 605-618.
- Lockhart, R. (1997). *Lecture 27: Variable Selection Methods*. Stat 350: Simon Fraser University.
- Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemotherapy Reports*, 50(8), pp. 163–170.
- Martin, J.B. & Douglas, G.A. (2004). Statistics Notes: The Log rank test. *British Medical Journal*. 328(7447), pp. 1073
- Martin, W.E., Swartz-Kulstad, J.L. & Madson, M. (1999). Psychosocial Factors That Predict the College Adjustment of First-Year Undergraduate Students: Implications for College Counsellors. *Journal of College Counselling*, 2(2), pp. 121-133.
- Marvasti, S.K., Rimaz, S., Abolghaesemi, J. & Heydari, I. (2017). Comparing of Cox Model and Parametric Models in Analysis of Effective Factors on Event Time of Neuropathy in Patients with Type 2 Diabetes. *Journal of Research in Medical Sciences*, 22(115), doi: 10.4103/jrms.JRMS\_6\_17.
- Mat, U., Buniyamin, N., Arsad, P.M. & Kassim, R.A. (2014). An Overview of using Academic Analytics to Predict and Improve Students' Achievement: A Proposed Proactive Intelligent Intervention. In *2013 IEEE 5th International Conference on Engineering Education: Aligning Engineering Education with Industrial Needs for Nation Development, ICEED 2013*, pp. 126-130.
- Meier, P. (1975). *Estimation of a Distribution Functions from Incomplete Observations in Perspectives in Probability and Statistics*. London: Academic Press.
- Menotti, A, Puddu, P.E., Loanni, M., Kromhou, D., Blackburn, H. & Nissinen, A. (2003). Twenty Five-Year Coronary Mortality Trends in the Seven Countries Study using the Accelerated Failure Time Model. *European Journal Epidemiology*, 18(2), pp. 113-122.
- Miller, L.D., Soh, L.K., Samal, A., Kupzyk, K. & Nugent, G. (2015). A Comparison of Educational Statistics and Data Mining Approaches to Identify Characteristics that Impact Online Learning. *Journal of Educational Data Mining*, 7(3), pp. 117-150.



- Min, Y., Zhang, G., Long, R.A., Anderson, T.J. & Ohland, M.W. (2011). Non-parametric Survival Analysis of the Loss Rate of Undergraduate Engineering Students. *Journal of Engineering Education*, 100(2), pp. 349-373.
- Minh, H.P. (2014). Survival Analysis: Breast Cancer. *Undergraduate Journal of Mathematical Modelling: One + Two*. 6(4), pp. 1-20.
- Morita, J.G., Lee, T.W. & Mowday, R.T. (1993). The Regression-Analog to Survival Analysis: A Selected Application. *Academy of Management Journal*, 36(6), pp. 1430–1464.
- Murray, S. & Tsatis, A.A. (1996). Non-parametric Survival Estimation using Prognostic Longitudinal Covariates. *Biometrics*, 52(1), pp. 137–151.
- Murtaugh, P.A., Burns, L.D. & Schuster, J. (1999). Predicting the Retention of University Students. *Research in Higher Education*, 40(3), pp. 355-371.
- Nardi, A. & Schemper, M. (2003). Comparing Cox and Parametric Models in Clinical Studies. *Statistics in Medicine*, 22(23), pp. 3597-3610.
- Nilsson, M. (2011). A Survival Analysis of Fixation Times in Reading. *Proceeding CMCL '11 Proceedings of the Second Workshop on Cognitive Modelling and Computational Linguistic*. Uppsala University. pp. 107-115.
- Orbe, J., Ferreira, E. & Nunez-Anton, V. (2002). Comparing Proportional Hazards and Accelerated Failure Time Models for Survival Analysis. *Statistics in Medicine*, 21(22), pp. 3493–3510.
- Patel, K., Kay, R. & Rowell, L. (2006). Comparing Proportional Hazards and Accelerated Failure Time Models: An Application in Influenza. *Pharmaceutical Statistics*, 5(3), pp. 213–224.
- Paura, L. & Arhipova, I. (2014). Cause Analysis of Student's Dropout Rate in Higher Education Study Program. *Procedia-Social and Behavioral Sciences*, 109(2014), pp. 1282 – 1286.
- Persson, I. & Khamis, H. (2005). Bias of the Cox Model Hazard Ratio. *Journal of Modern Applied Statistical Methods*, 4(1), pp. 90-99.
- Pourhoseingholi, M.A., Hajizadeh, E., Dehkordi, B.M., Safaee, A., Abadi, A. & Zali, M.R. (2007). Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. *Asian Pacific Journal of Cancer Prevention*, 8(3), pp. 412–416.

- Pratschke, J., Haase, T., Comber, H., Sharp, L., Cancela, M.C. & Johnson, H. (2016). Mechanisms and Mediation in Survival Analysis: Towards an Integrated Analytical Framework. *BMC Medical Research Methodology*, 16(27), pp. 1-13.
- Prel, J.B., Hommel, G., Rohrig, B. & Blettner, M. (2009). Confidence Interval or P-Value? *Deutsches Arzteblatt International*, 106(19), pp. 335–339.
- Prinja, S., Gupta, N. & Verma, R. (2010). Censoring in Clinical Trials: Review of Survival Analysis Techniques. *Indian Journal of Community Medicine*, 35(2), pp. 217– 221.
- Quantin C., Moreau, T., Asselain B., Maccario J. & Lellouch, J. (1996). A Regression Survival Model for Testing the Proportional Hazards Hypothesis. *Biometrics*, 52(3), pp. 874-885.
- Rasul, S. & Bukhsh, Q. (2011). A Study of Factors Affecting Students Performance in Examination at University Level. *Procedia Social and Behavioral Sciences*, 15(2011), pp. 2042-2047.
- Reibnegger, G. (2012). Modeling Time in Medical Education Research: The Potential of New Parametric Methods of Survival Analysis. *Scientific Research*, 3(special issue), pp. 916-922.
- Rich, J.T., Neely, J.G.N., Paniello, R.C., Voelker, C.C.J., Nussenbaum, B. & Wang, E.W. (2010). A Practical Guide to Understanding Kaplan-Meier Curves. *Otolaryngol Head Neck Surgery Journal*, 143(3), pp. 331-336.
- Ritesh, S. & Keshab, M. (2011). Survival Analysis in Clinical Trials: Basics and Must Know Areas. *Perspective in Clinical Research*, 2(4), pp. 145-148.
- Romero, C. & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on System Man and Cybernetics Part C (Applications and Reviews)*. 40(6), pp. 601–618.
- Saikia, R. & Barman, M.P. (2016). Comparing Cox Proportional Hazard Model and Parametric Counterpart in the Analysis of Esophagus Cancer Patient Data, *Journal of Mathematics*, 12(5), pp. 16-21.
- Saikia, R. & Barman, M.P. (2017). A Review on Accelerated Failure Time Models. *International Journal of Statistics and Systems*, 12(2), pp. 311-312.
- Schmidt, P. & Witte, A. (1991). Survival analysis: A survey. *Journal of Quantitative Criminology*. 7(1), pp. 59-68.

- Schoenfeld, D. (1982). Partial Residuals for the Proportional Hazards Regression Model. *Biometrika*, 69(1982), pp. 239 – 241.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), pp. 461- 464
- Shahzadi, E. & Ahmad, Z. (2011). A Study on Academic Performance of University Students. *Conference: 8<sup>th</sup> International Conference on Recent Advances in Statistics, At Lahore, Pakistan. February 8-9*, pp. 255-268.
- Spruance, S.L., Reid, J.E., Grace, M. & Samore, M. (2004). Hazard Ratio in Clinical Trials. *Antimicrobial Agents and Chemotherapy*, 48(8), pp. 2787-2792.
- Steinmayr, R., Meissner, A., Weidinger, A.F. & Wirthwein, L. (2014). *Academic Achievement*. United Kingdom: Oxford University Press.
- Swindell, W.R. (2009). Accelerated Failure Time Models Provide a Useful Statistical Framework for Aging Research. *Experimental Gerontology Journal*, 44(3), pp. 190-200.
- Symonds, M.R.E. & Moussalli, A. (2011). A Brief Guide to Model Selection, Multimodel Inference and Model Averaging in Behavioral Ecology using Akaike's Information Criterion (AIC). *Behavioral Ecology and Socio-biology*, 65(1), pp. 13-21.
- Tarone, R.E. & Ware, J. (1977). On Distribution-Free Tests for Equality of Survival Distributions. *Biometrika*, 64(1), pp. 156-160.
- Teshnizi, S.H. & Ayatollahi, S.M.T. (2017). Comparison of Cox Regression and Parametric Models: Application for Assessment of Survival of Pediatric Cases of Acute Leukemia in Southern Iran. *Asian Pacific Journal of Cancer Prevention*, 18(4), pp. 981-985.
- Therneau, T.M., Grambsch, P.M. & Fleming, T.R. (1990). Martingale-based Residuals for Survival Models, *Biometrika*. 7, pp. 147 – 160.
- Therneau, T.M. & Grambsch, P.M. (2000). *Modelling Survival Data: Extending the Cox Model*. New York: Springer.
- Tianqi, H. & Ganges, T.W. (1995). A Discrete-Time Survival Analysis of the Education Path of Specially Admitted Students. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco.
- Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. 2<sup>nd</sup> ed. Chicago: University of Chicago Press.

- Tinto, V. (1997). Colleges as Communities: Exploring the Educational Character of Student Persistence. *Journal of Higher Education*, 68(6), pp. 599-623.
- Tinto, V (2006). Research & Practice of Students' Retention: What next? *Journal of College Students Retention*, 8(1), pp. 1-19.
- Vartanian, T.P. (2011). *Secondary Data Analysis*. New York: Oxford University Press.
- Wei, L.J. (1992). The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis. *Statistics in Medicine*, 11(14-15), pp. 1871-1879.
- Weibull, W.A. (1951). A Statistical Theory of the Strength of Materials. *Ingeniors Vetenskaps Akademiens, Handlingar*, 151(3). pp. 293-297.
- Wu, J. (2014). One Sample Log-Rank test. *Journal of Biometric & Biostatistics* 5(210), doi:10.4172/2155-6180.1000210
- Yamaguchi, K. (1994). Some Accelerated Failure-Time Regression Models Derived from Diffusion Process Models. *Sociology Methodology Journal*, 24, pp. 267-300.
- Yang & Faxian (2000). Using Survival Analysis To Analyse and Predict Students' Achievement from Their Status of Development Study. *Paper Presented at the Annual meeting of The Association for International Research (40<sup>th</sup>, Cincinnati oh, May 21-24, 2000)*.
- Zare, A., Hosseini, M., Mahmoodi, M., Mohammad, K., Zeraati, H. & Holakouie Naieni, K. (2015). A Comparison between Accelerated Failure-time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients. *Iranian Journal of Public Health*, 44(8), pp. 1095–1102.
- Zhao, G. (2008). *Non-parametric Survival Analysis of Censored Data with Possible Violation of Method Assumptions*. University of North Carolina, Greensboro. Master's Thesis.
- Zhu, X., Zhou, X., Zhang, Y., Sun, X., Liu, H. & Zhang, Y. (2017). Reporting and Methodological Quality of Survival Analysis in Articles Published in Chinese Oncology Journals. *Medicine (Baltimore)*, 96(50), pp. 1-7.
- Zuaini, I., Aziah, A.M. & Rosliza, M.Z. (2008). Factor that Determine Matriculation Students in Account Subjects. *Malaysian Journal of Learning and Instruction*, 5, pp. 99-115.

Zuilkowski, S.S. & Jukes, M.C.H. (2014). Early Childhood Malaria Prevention and Children's Pattern of School Leaving in The Gambia. *British Journal of Educational Psychology*, 84, pp. 483-501.

